Probability and Statistics for Ensemble Forecasting

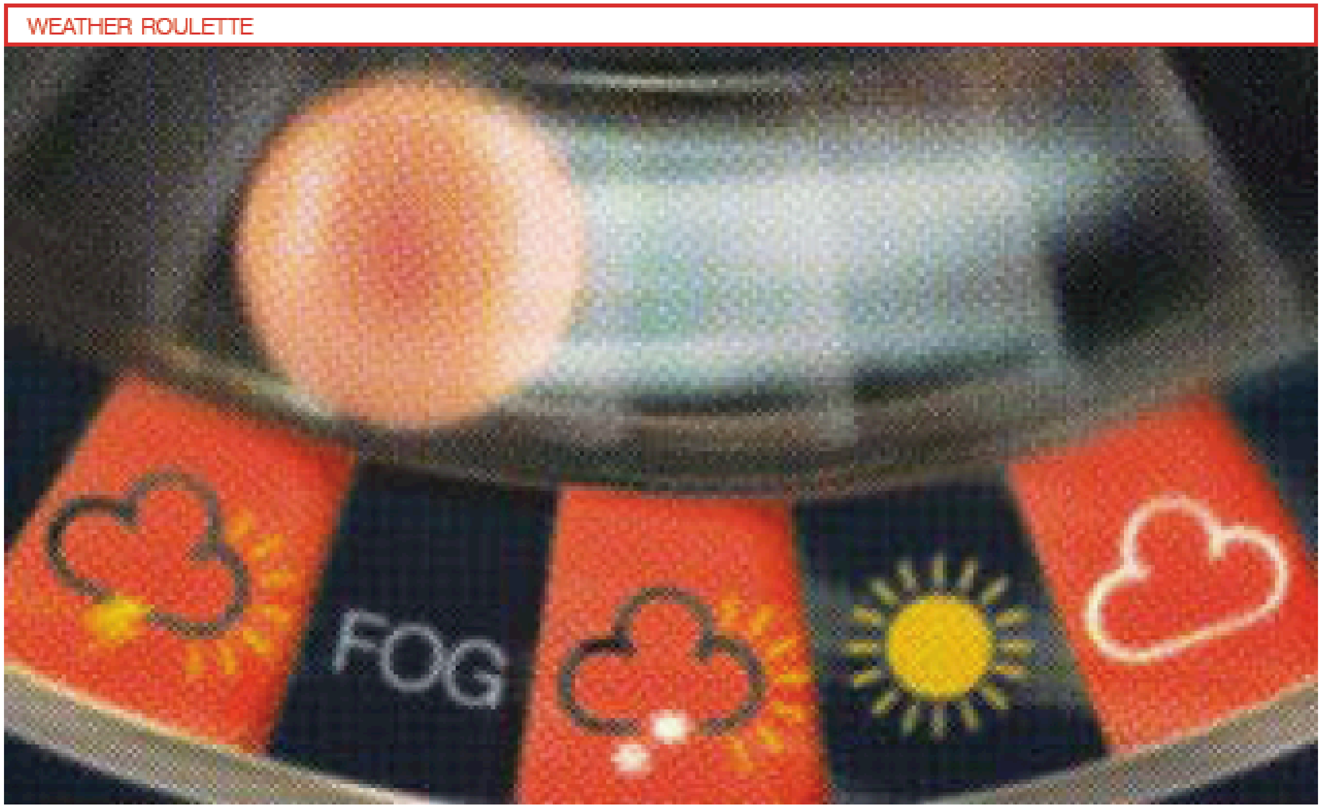
Tom Hamill (NOAA/ESRL, Boulder)
and
Jim Hansen (Navy/NRL, Monterey)

(borrows heavily from Dan Wilks'
Statistical Methods in the Atmospheric Sciences text)

Probability and statistics

- **Probability**: a formalism for expressing uncertainty quantitatively.
- **Statistics**: the science pertaining to the collection, analysis, interpretation or explanation, and presentation of data.
- **Goal**: get you comfortable with the terminology the other instructors will use.

Part 1: Probability

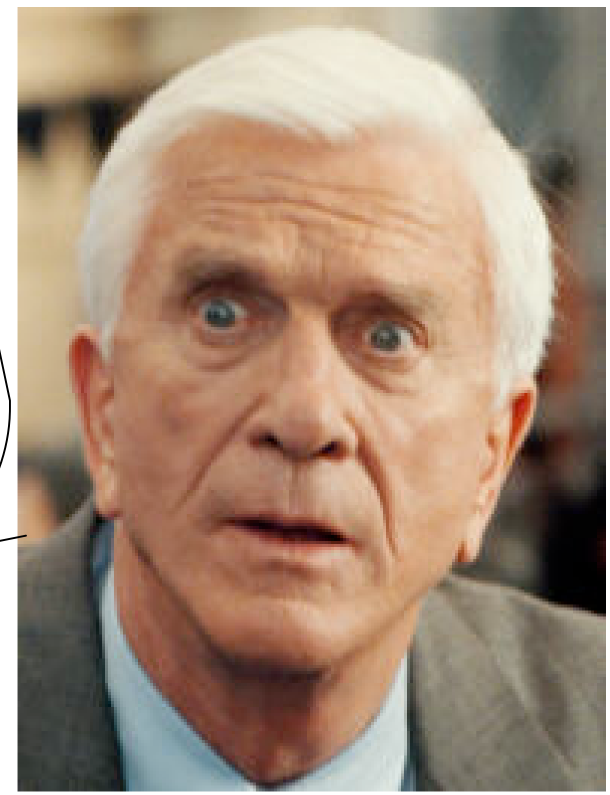


Weather is uncertain, so we use the language of uncertainty

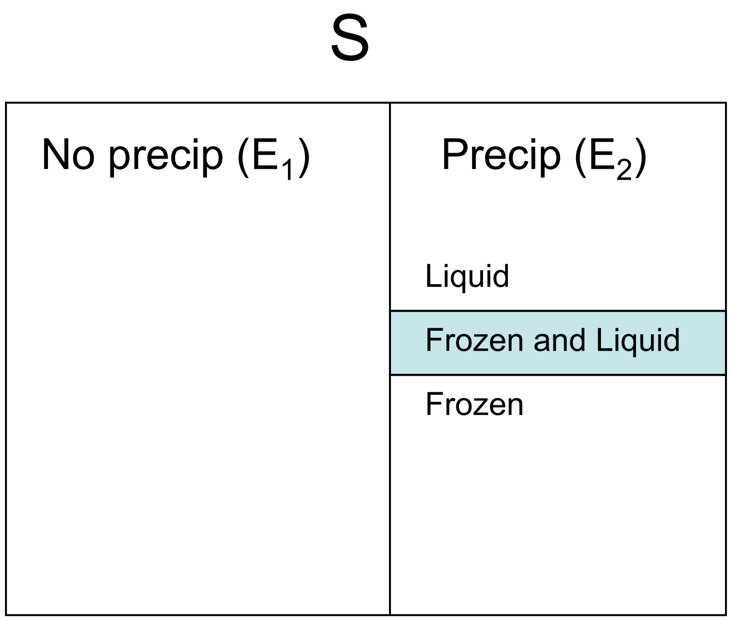
photo courtesy of Lenny Smith, Oxford U. and London School of Economics

Is probability (1) inherently confusing, or (2) a formal way of bamboozling and waffling?

"Doctors say that Nordberg has a 50/50 chance of living, though there's only a 10 percent chance of that."



Axioms of Probability



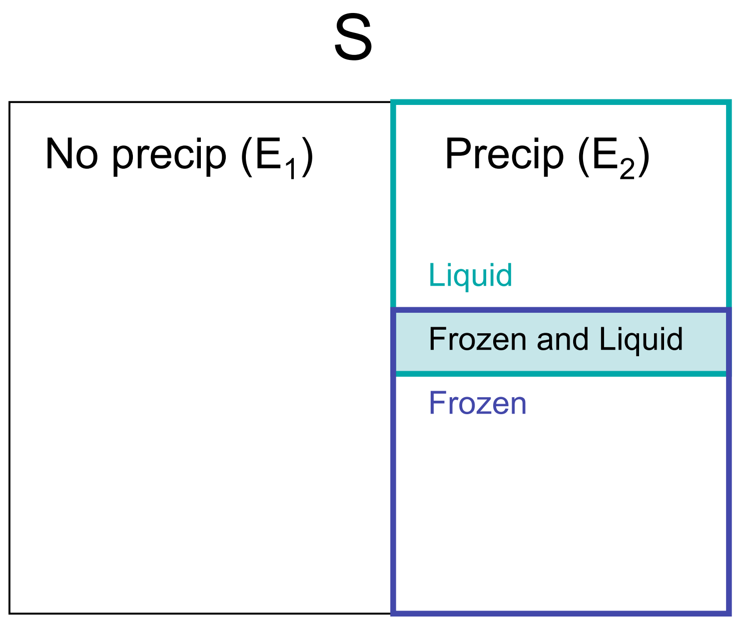
$0.0 \leq Pr(E_1) \leq 1.0$

$Pr(S) = 1.0$

$Pr(E_1) + Pr(E_2) = 1.0$

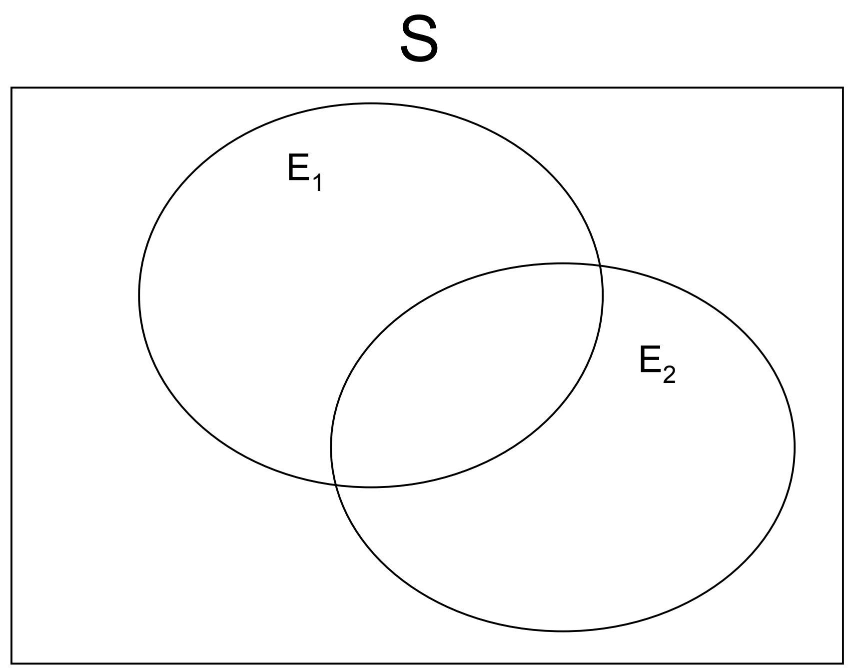
S is the “sample space.” E_1 and E_2 are “mutually exclusive” and “collectively exhaustive” events that fill the sample space.

Union of Events

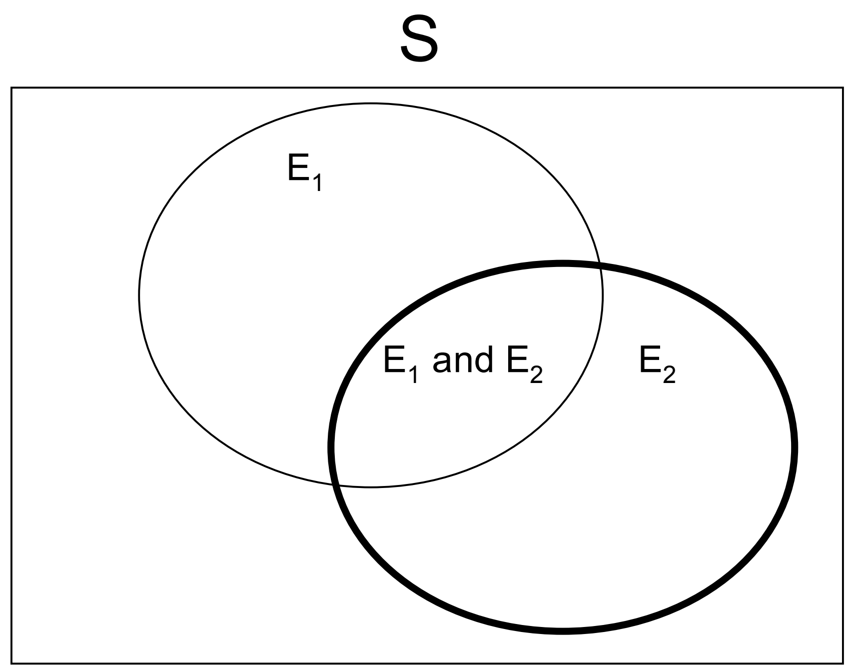


Pr(Liquid) and Pr(Frozen) = Pr(Liquid) + Pr(Frozen) - Pr(Frozen and Liquid)

Conditional Probability



Conditional Probability



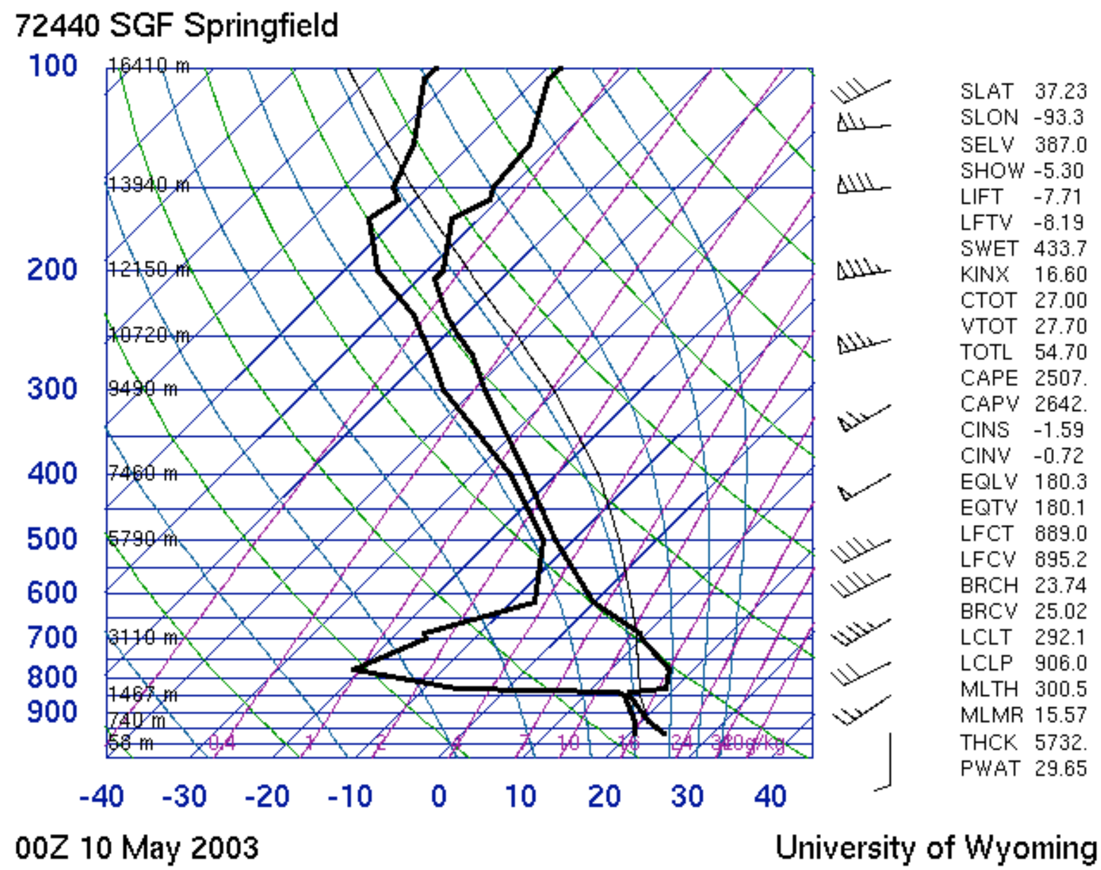
$$\Pr(E_1 \mid E_2) = \Pr(E_1 \text{ given that } E_2 \text{ has occurred})$$
$$= \Pr(E_1 \text{ and } E_2) / \Pr(E_2)$$

narrow the playing field … consider only the subset where E_2 has occurred

Example: “loaded gun” sounding

P(tornado in SW MO) = 0.02

unconditional probability of a tornado is small; most likely it will be impossible to break through the capping inversion.



P(tornado | thunderstorm) = 0.35

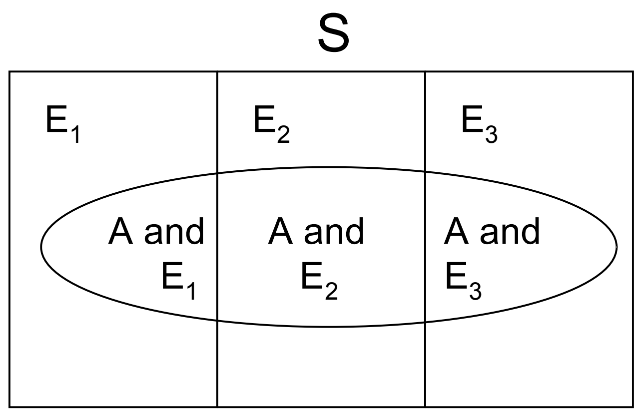
if penetrative convection does happen, the large instability and shear increase the probability that the thunderstorm will produce a tornado.

Independence

- E_1 and E_2 are independent if and only if $\Pr(E_1 \text{ and } E_2) = \Pr(E_1) \times \Pr(E_2)$

Probability of two sixes = $1/6 \times 1/6 = 1/36$

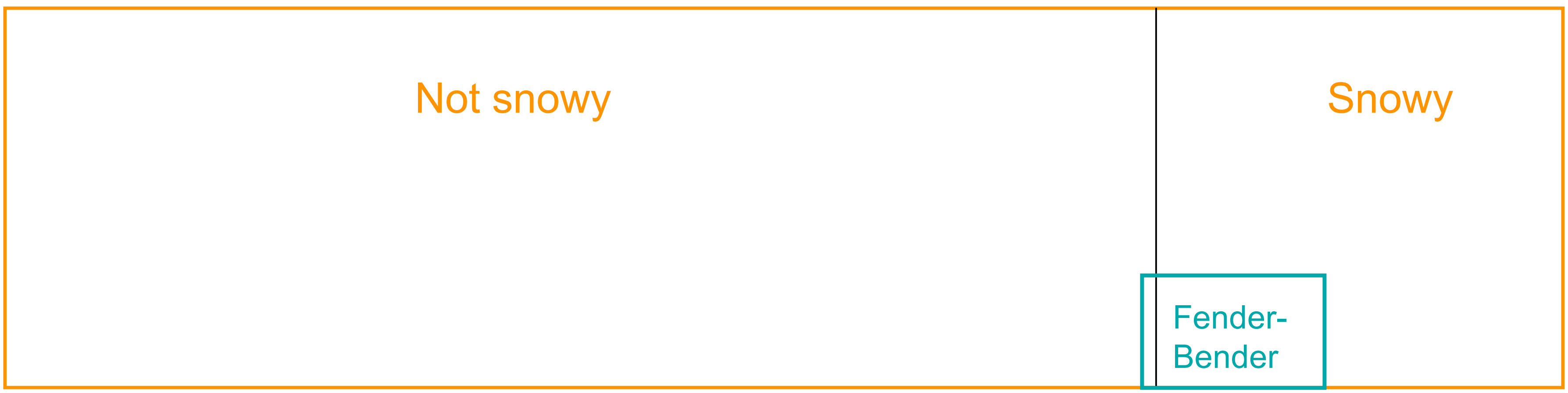
Law of total probability



$$\Pr(A) = \sum_{i=1}^{3} \Pr(A \mid E_i) \Pr(E_i)$$

Overall “unconditional” probability can be computed summing / integrating the weighted conditional probabilities

Law of total probability: driving example

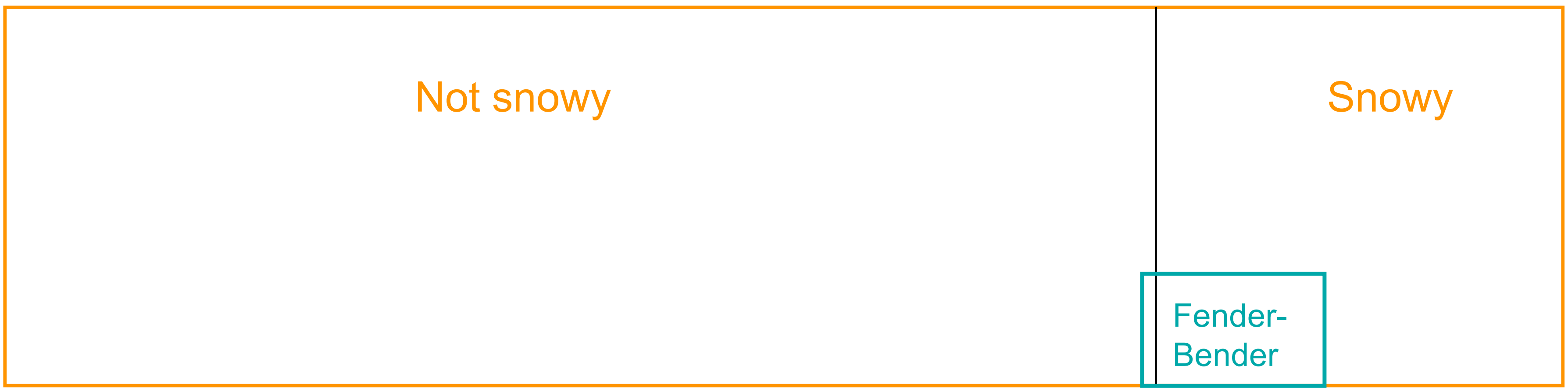


P(Fender-Bender) = P(Fender-Bender| Not Snowy) P(Not Snowy) +
P(Fender-Bender| Snowy) P(Snowy)

= 0.01 x 0.75 +
0.10 x 0.25

= 0.0325

Law of total probability: driving example



P(Fender-Bender) = P(Fender-Bender| Not Snowy) P(Not Snowy) +
P(Fender-Bender| Snowy) P(Snowy)

$$= 0.01 \times 0.75 + 0.10 \times 0.25$$

= 0.0325 (I’m an excellent driver)

Discrete vs. Continuous Probability

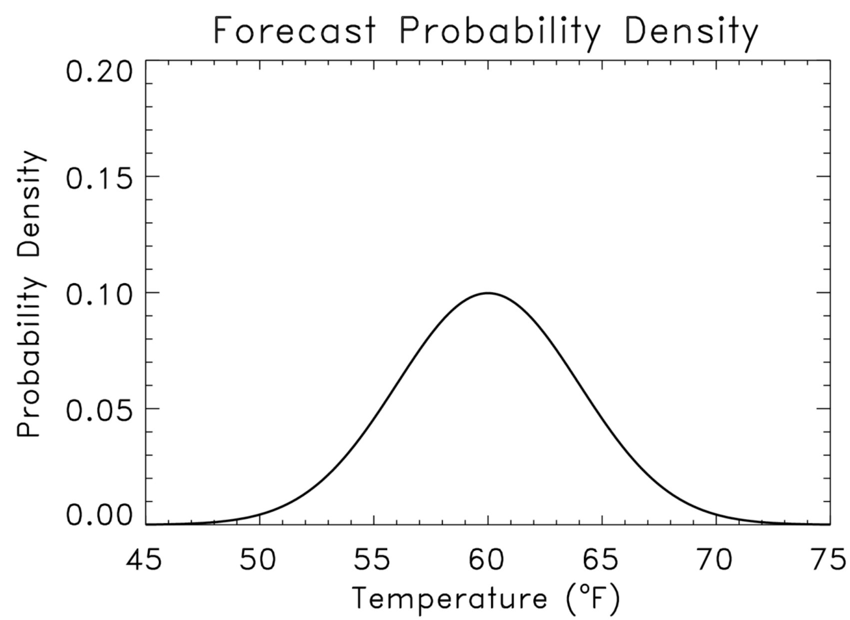
- **Discrete**: limited number of possible outcomes

- **Continuous**: unlimited number of outcomes.

P(T=60.0) not meaningful;

probability density
expressed relative likelihood
of being *near* a particular value;
and probability density follows
other probability axioms, e.g.,

$$\int_{t=0K}^{\infty} P(t)dt = 1.0$$



Discrete "parametric" probability distributions: the binomial distribution

$$\Pr(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$$

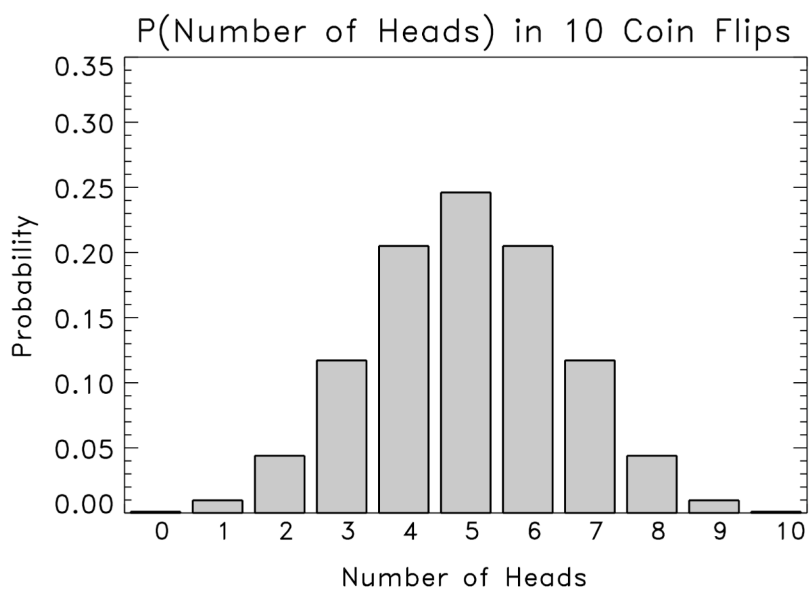
$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

X is random variable

x is a specific number

N is the number of trials

p is the event probability



Discrete "parametric" probability distributions: the binomial distribution

$$\Pr(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$$

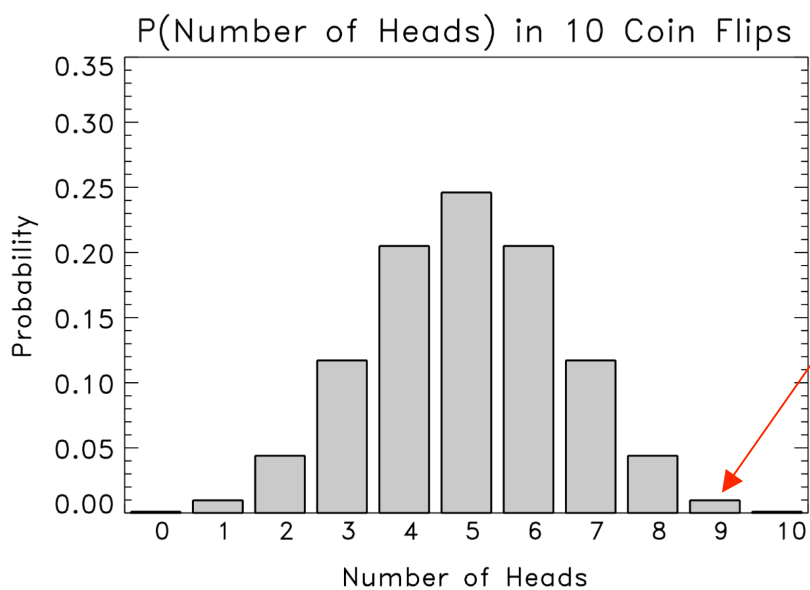
$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

X is random variable

x is a specific number

N is the number of trials

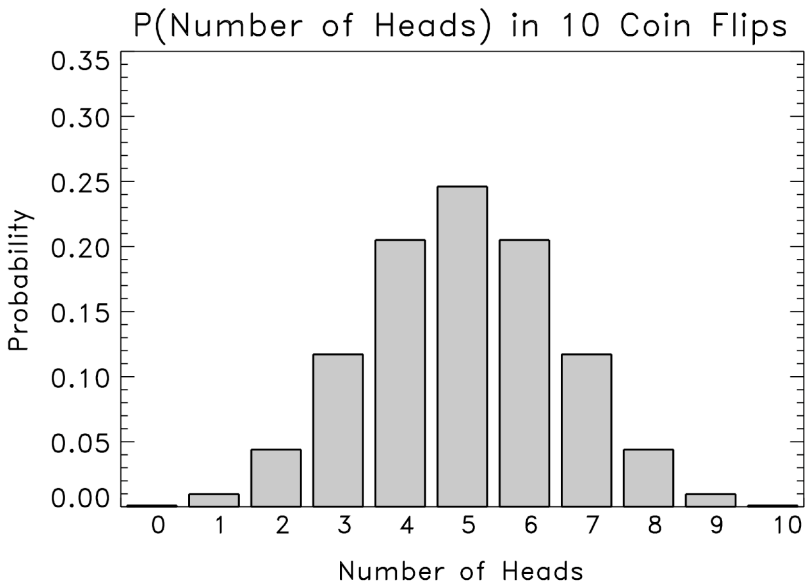
p is the event probability



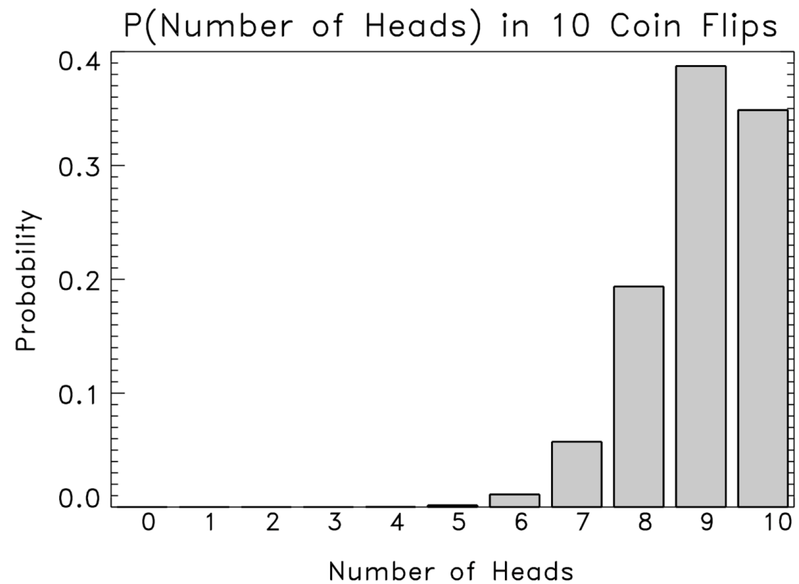
analogy: you might forecast 50% probability of rain, and rain may happen 9/10 times. That can happen, though it's unlikely.

Binomial distributions

p = 0.5



p = 0.9



…though perhaps p = 0.9 would have been a more appropriate choice.

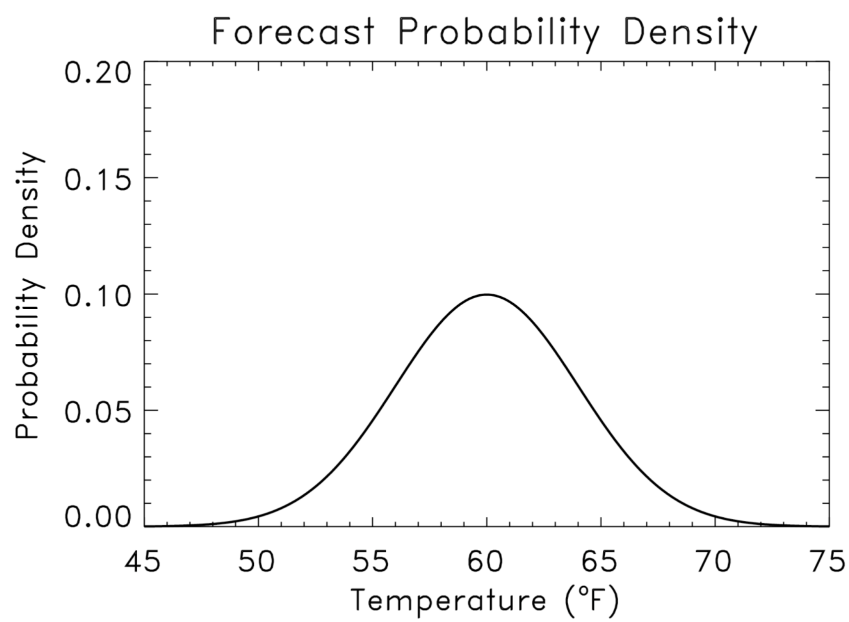
Continuous parametric probability distributions: the Normal distribution

- Also called "Gaussian" or 'the bell-shaped curve"

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$$

- $f(x)$ is the "probability density"
- μ is the "mean"
- σ is the "standard deviation"

μ = 60.0, σ = 4.0



Continuous parametric probability distributions: the Normal distribution

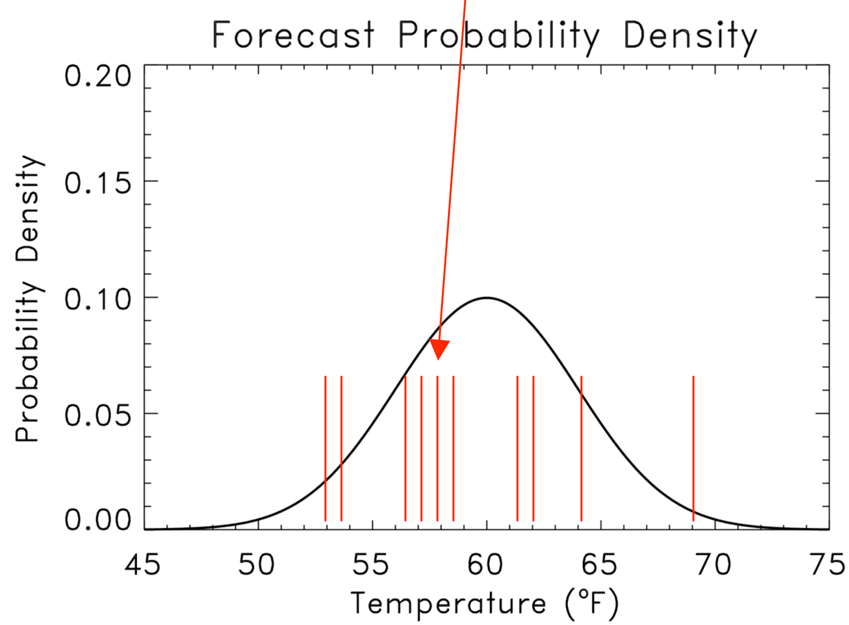
- Also called "Gaussian" or 'the bell-shaped curve"

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$$

this ensemble might be a random sample from a smooth distribution like this

- $f(x)$ is the "probability density" function, or PDF
- μ is the "mean"
- σ is the "standard deviation"

$\mu = 60.0,\ \sigma = 4.0$

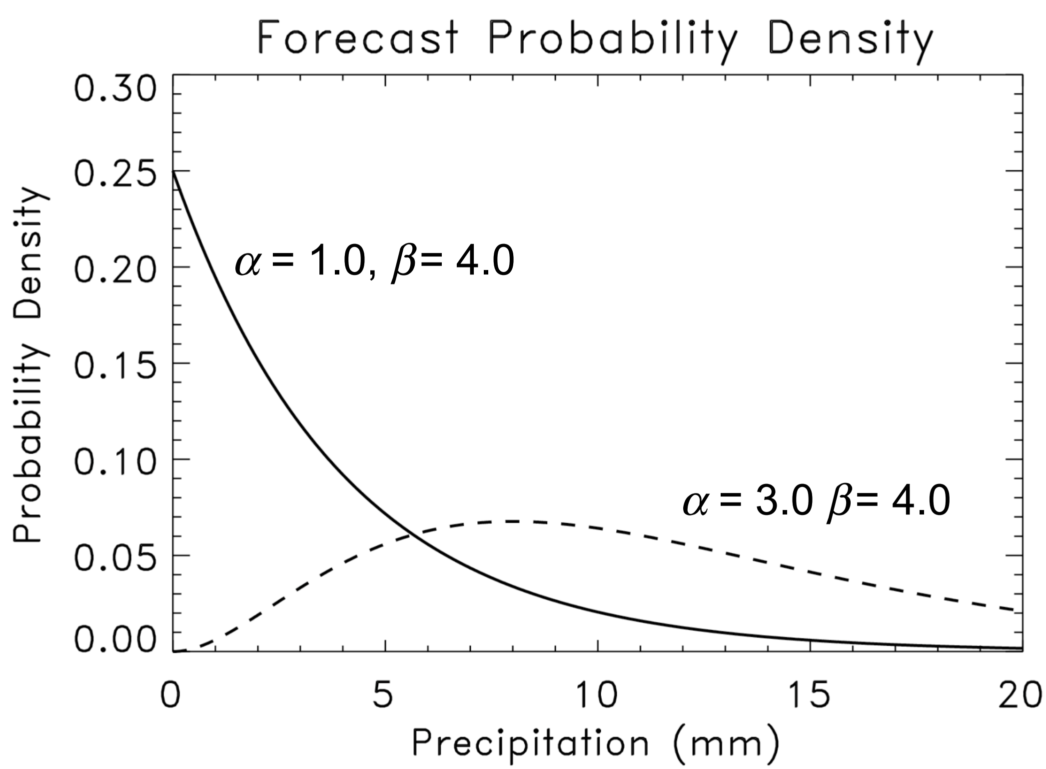


The gamma distribution

$$f(x) = \frac{\left(x/\beta\right)^{\alpha-1} \exp\left(-x/\beta\right)}{\beta\Gamma(\alpha)}, \qquad x, \alpha, \beta > 0.0$$

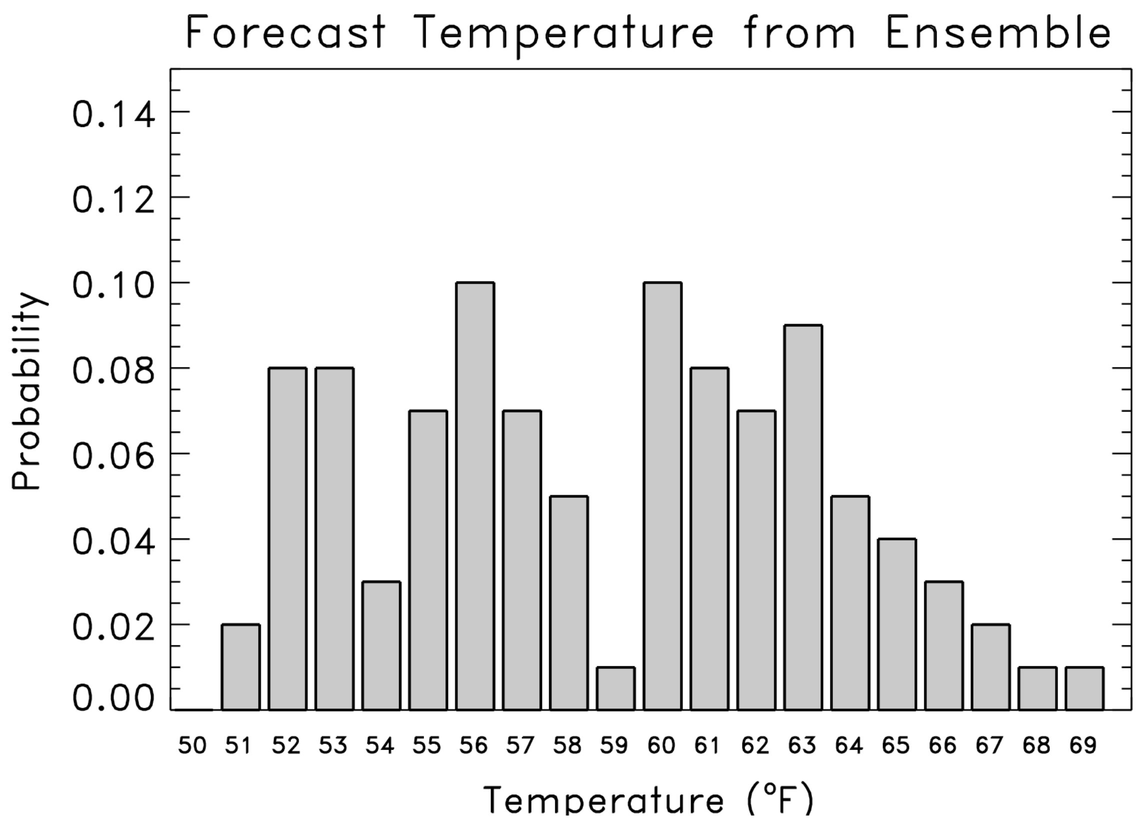
α = shape parameter
β = scale parameter

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$



"Empirical" probability distributions

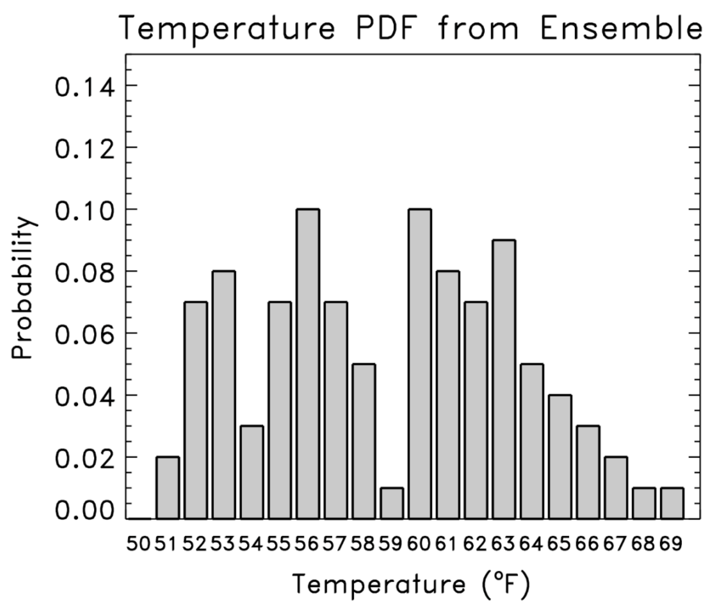
distribution derived from the data itself

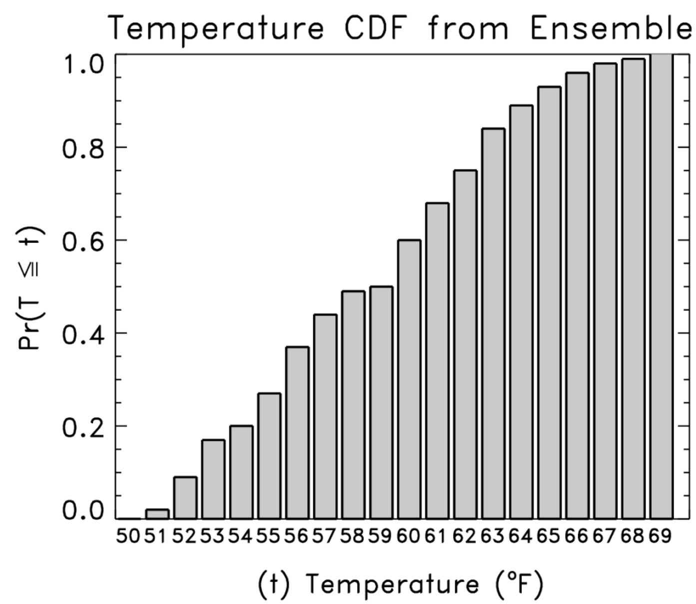


Cumulative Distribution Function (CDF)

- $F(t) = \Pr\{T \le t\}$

  where T is the random variable, t is some specified threshold.





Statistics

- **Definition**: “the science pertaining to the collection, analysis, interpretation or explanation, and presentation of data.”

- **Goal**: make sure you understand terminology that we’ll be using (mean, standard deviation, correlation, covariance, etc.)

Measures of “location”

- T = [50, 51, 53, 54, 54, 57, 59, 63, 65, 66, 84]
  (n=11)
- Measure the centrality of this data set in some fashion.
- Mean (also called average, or 1st moment); minimizes RMS error:

$$\bar{T} = \frac{\sum_{i=1}^{n} T_i}{n} = 59.63$$

- Median: central value of the sample, here = 57. Less affected by the 84 “outlier.” Minimizes mean absolute error.

Measures of spread

- T = [50, 51, 53, 54, 54, 57, 59, 63, 65, 66, 84]

- Standard Deviation of sample:

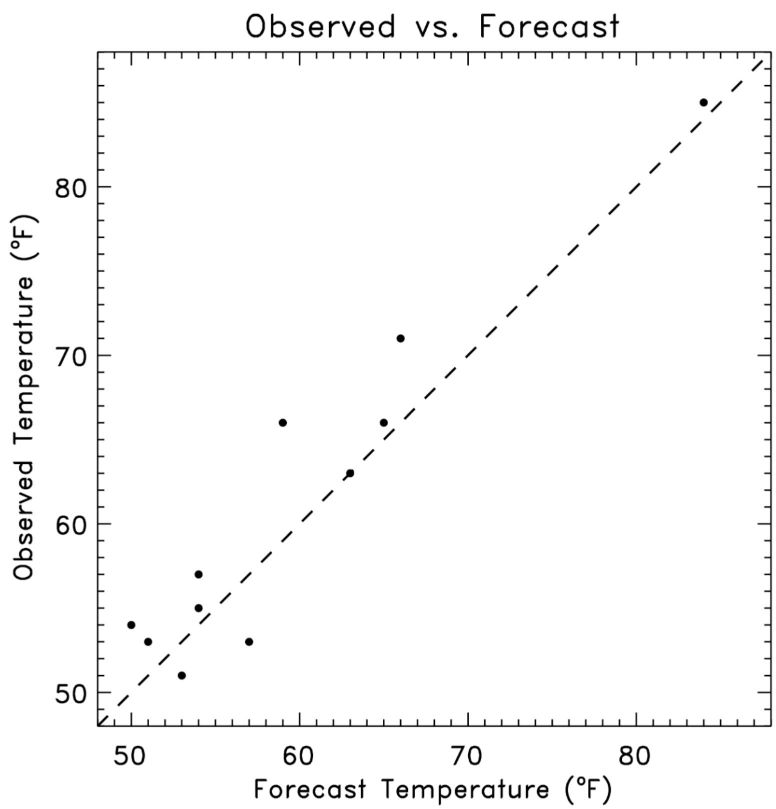
  (variance is the square of this)

$$s = \left[\frac{\sum_{i=1}^{n} \left(T_i - \bar{T} \right)^2}{n-1} \right]^{1/2} = 9.78$$

- IQR (Interquartile Range) = $q_{0.75}$ - $q_{0.25}$ = 65 - 53 = 12
  where $q_{0.75}$ is the 75th percentile (quantile) of the distribution and $q_{0.25}$ is the 25th percentile.

Measures of association

- T_f = [50, 51, 53, 54, 54, 57, 59, 63, 65, 66, 84]
- T_o = [54, 53, 51, 57, 55, 53, 66, 63, 66, 71, 87]



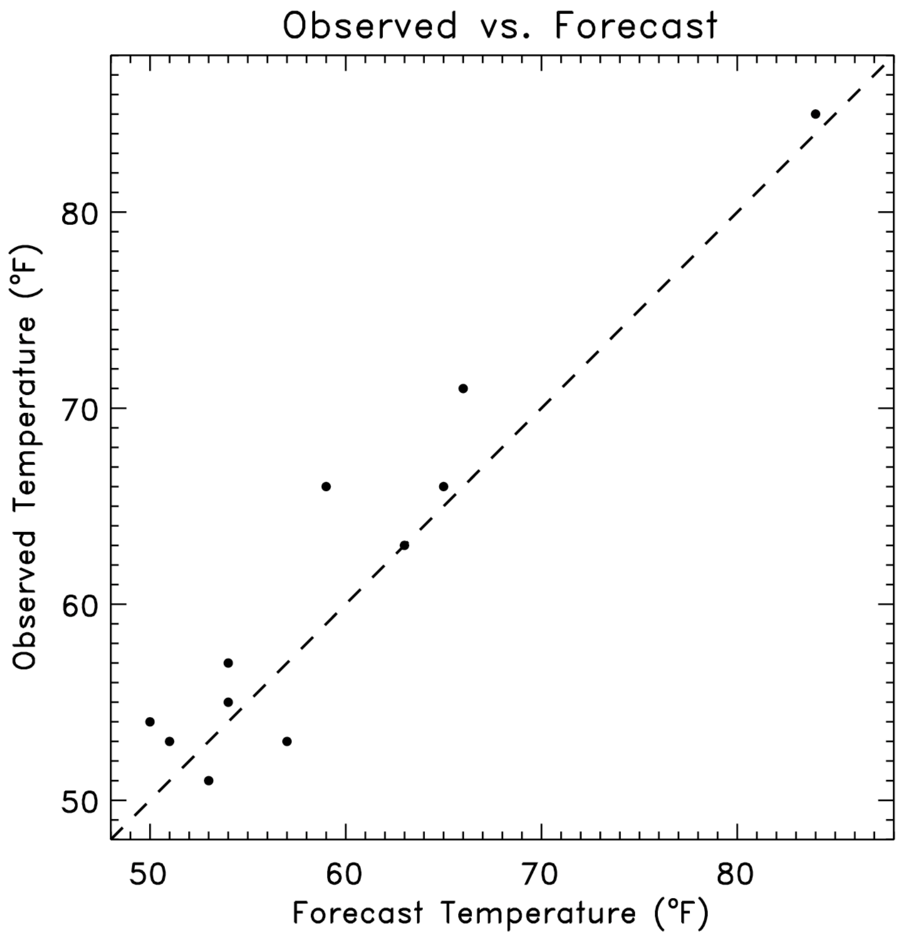
Measures of association

- Pearson (ordinary) correlation:

$$r_{xy} = \frac{Cov(x,y)}{s_x s_y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}\left[(x_i - \overline{x})(y_i - \overline{y})\right]}{\left\{\frac{1}{n-1}\sum_{i=1}^{n}\left[(x_i - \overline{x})^2\right]^{1/2}\right\}\left\{\frac{1}{n-1}\sum_{i=1}^{n}\left[(y_i - \overline{y})^2\right]^{1/2}\right\}}$$

$$= \frac{\sum_{i=1}^{n}[x_i' \, y_i']}{\left(\sum_{i=1}^{n}[x_i']^2\right)^{1/2}\left(\sum_{i=1}^{n}[y_i']^2\right)^{1/2}}$$

Correlation, mean, standard deviation



$$r = 0.953$$

$$\bar{T}_f = 59.63$$

$$s\left(T_f\right) = 9.78$$

$$\bar{T}_o = 61.27$$

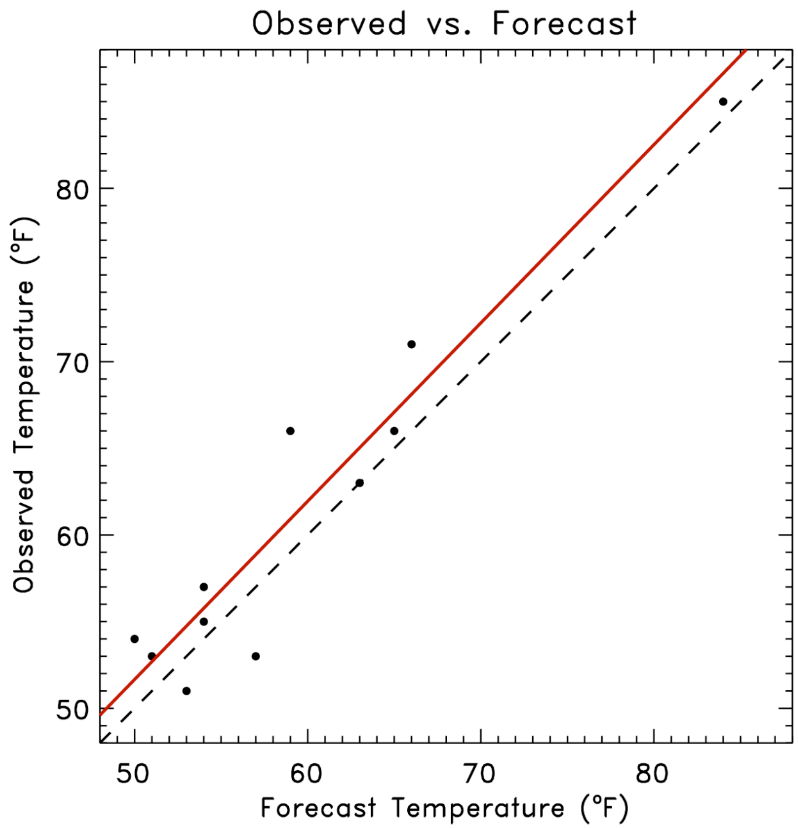
$$s\left(T_o\right) = 10.28$$

Regression

Find the equation that minimizes the squared difference between forecasts and observations.

$$T_o^{pred} = 1.478 + 1.006 * T^f$$

Methods like this used to statistically adjust weather forecasts.

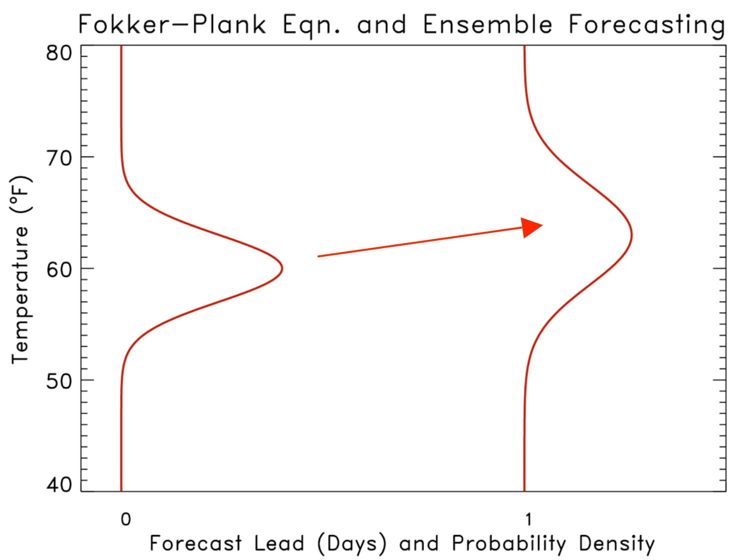


Connection between ensemble forecasts and PDFs

(there is a theory behind ensemble forecasting!)

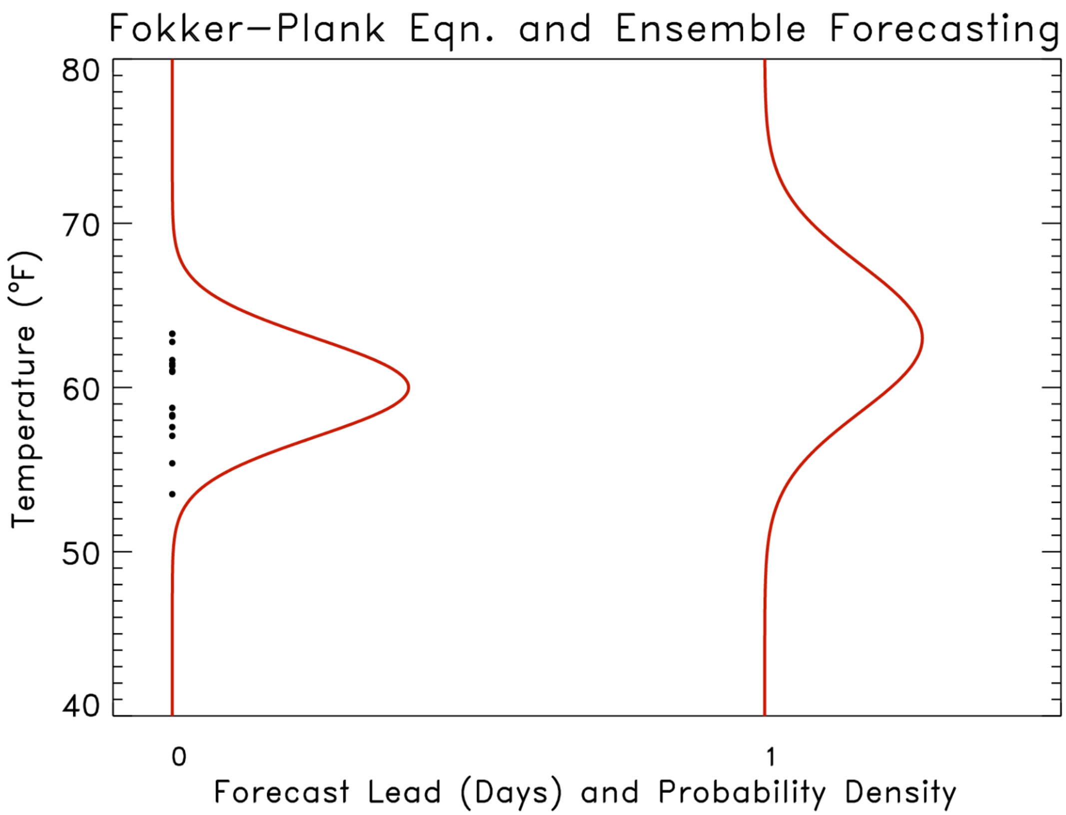
Fokker-Planck equation to describe evolution of forecast PDF

$$\frac{\partial P\left(\mathbf{x}_t^t\right)}{\partial t} = -\nabla \bullet \left[\underbrace{M\left(\mathbf{x}_t^t\right)}_{\text{errors due to chaos}} P\left(\mathbf{x}_t^t\right)\right] + \sum_{i,j} \frac{\partial^2}{\partial \mathbf{x}_{t(i)}^t \partial \mathbf{x}_{t(j)}^t} \underbrace{\left(\frac{G\mathbf{Q}_t G^T}{2}\right)_{i,j}}_{\text{errors due to the model}} P\left(\mathbf{x}_t^t\right)$$



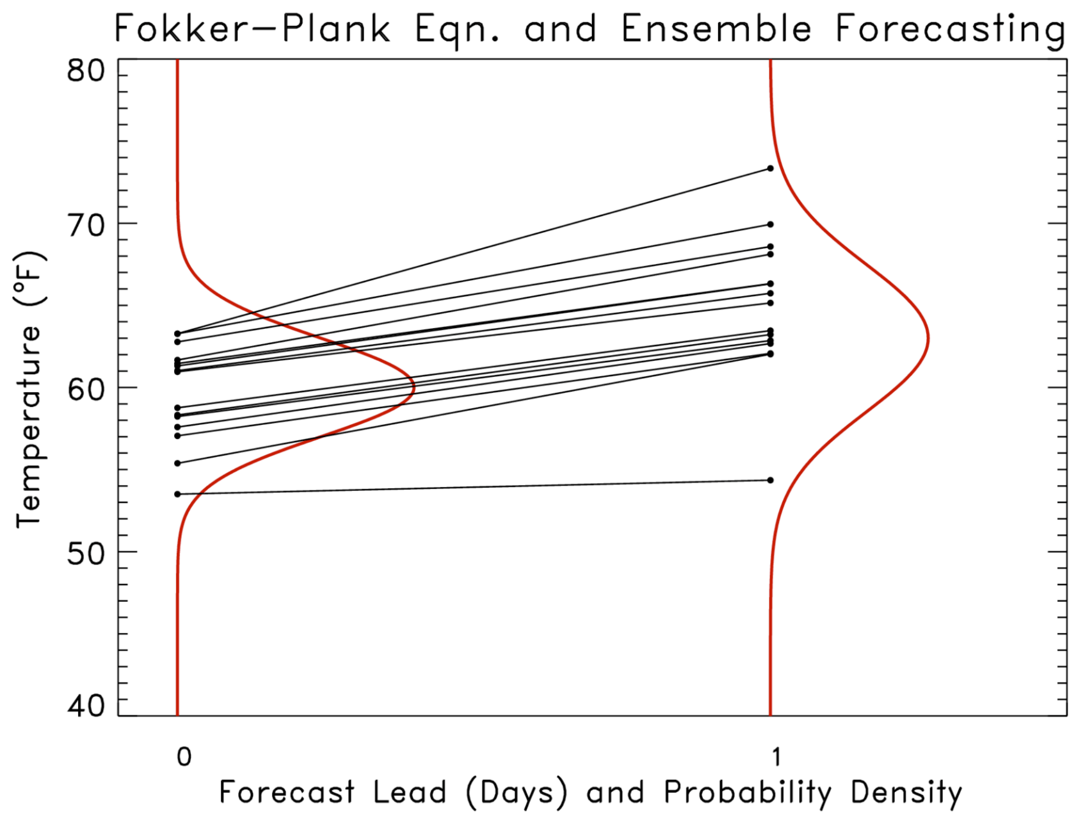
(in reality, we never can get the pdf shown on day 1)

Connection, cont'd



In ensemble forecasting (ideally), we sample the initial pdf, and…

Connection, cont'd



In ensemble forecasting (ideally), we sample the initial PDF, and evolve each initial condition forward with the forecast model(s) to randomly sample the day-1 PDF

Questions?

Baye's Rule

$$\Pr(A \; and \; E_1) = \Pr(A \mid E_1)\Pr(E_1)$$
$$= \Pr(E_1 \mid A)\Pr(A)$$

combine 2 right-hand sides and rearrange

$$\Pr(E_1 \mid A) = \frac{\Pr(A \mid E_1)\Pr(E_1)}{\Pr(A)} = \frac{\Pr(A \mid E_1)\Pr(E_1)}{\sum_{j=1}^{J}\Pr(A \mid E_j)\Pr(E_j)}$$